

What is claimed is:

1 1. An apparatus for compressing a plurality of  
2 structured documents having a common data structure, said  
3 apparatus comprising:

4 a tag list obtaining unit for obtaining a single  
5 tag list, common to said plural structured documents,  
6 that lists tags in the order of appearance;

7 a structured document compressing unit for  
8 generating a plurality of compressed documents in which  
9 tags in individual said plural structured documents are  
10 replaced with predetermined delimiter codes; and

11 an outputting unit for outputting said single  
12 tag list, which is obtained by said tag list obtaining  
13 unit, and also said plurality of compressed documents,  
14 which are generated individually from said plural  
15 structured documents by said structured document  
16 compressing unit, in correspondence with one another.

1 2. A structured document compressing apparatus  
2 according to claim 1, wherein said structured document  
3 compressing unit further comprises:

4 a tag detecting unit for detecting each tag in  
5 individual said structured documents; and

6 a tag replacement unit for replacing said tag,  
7 detected by said tag detecting unit, with said  
8 predetermined delimiter code.

1 3. An apparatus for compressing a structured  
2 document, said apparatus comprising:  
3 a tag detecting unit for detecting each tag in  
4 said structured document; and  
5 a tag replacement unit for replacing said tag,  
6 detected by said tag detecting unit, with a predetermined  
7 delimiter code.

1 4. An apparatus for compressing a structured  
2 document, said apparatus comprising:  
3 a subdocument extracting unit for extracting a  
4 subdocument, which is a region sandwiched between a start  
5 tag and an end tag that have a predetermined element name,  
6 from said structured document;  
7 a tag detecting unit for detecting each tag in  
8 said subdocument extracted by said subdocument  
9 extracting unit; and  
10 a tag replacement unit for replacing said tag,  
11 detected by said tag detecting unit, with a predetermined  
12 delimiter code.

1 5. A structured document compressing apparatus  
2 according to claim 3, further comprising:  
3 an attribute-bearing-tag discriminating unit  
4 for discriminating whether or not said tag detected by  
5 said tag detecting unit is an attribute-bearing tag, which  
6 has an attribute value; and

7           an attribute-bearing-tag replacement unit for  
8 replacing said attribute-bearing tag, discriminated by  
9 said attribute-bearing-tag discriminating unit, with a  
10 set of the attribute value and a predetermined delimiter  
11 code.

1   6.       A structured document compressing apparatus  
2 according to claim 4, further comprising:

3           an attribute-bearing-tag discriminating unit  
4 for discriminating whether or not said tag detected by  
5 said tag detecting unit is an attribute-bearing tag, which  
6 has an attribute value; and

7           an attribute-bearing-tag replacement unit for  
8 replacing said attribute-bearing tag, discriminated by  
9 said attribute-bearing-tag discriminating unit, with a  
10 set of the attribute value and a predetermined delimiter  
11 code.

1   7.       A structured document compressing apparatus  
2 according to claim 3, further comprising:

3           a tag list holding unit for holding a tag list  
4 in which tags are listed in a predetermined order for  
5 definition of a predetermined data structure;

6           a tag rearranging unit for rearranging tags in  
7 said structured document before compressed, in the  
8 predetermined order according to the tag list held in  
9 said tag list holding unit; and

10           an omitted-tag supplementing unit for  
11 supplementing a tag omitted in said structured document  
12 according to said tag list held in said tag list holding  
13 unit.

1   8.       A structured document compressing apparatus  
2 according to claim 4, further comprising:

3           a tag list holding unit for holding a tag list  
4 in which tags are listed in a predetermined order for  
5 definition of a predetermined data structure;

6           a tag rearranging unit for rearranging tags in  
7 said structured document before compressed, in the  
8 predetermined order according to the tag list held in  
9 said tag list holding unit; and

10           an omitted-tag supplementing unit for  
11 supplementing a tag omitted in said structured document  
12 according to said tag list held in said tag list holding  
13 unit.

1   9.       A structured document compressing apparatus  
2 according to claim 5, further comprising:

3           a tag/attribute list holding unit for holding  
4 a tag/attribute list in which tags and an attribute name  
5 are listed in a predetermined order for the definition  
6 of a predetermined data structure;

7           a tag/attribute rearranging unit for rearranging  
8 tags and an attribute in the structured document to be

9 compressed, in the predetermined order according to the  
10 tag/attribute list held in said tag/attribute list  
11 holding unit; and  
12 an omitted tag/attribute supplementing unit for  
13 supplementing a tag and/or an attribute omitted in said  
14 structured document according to the tag/attribute list  
15 held in said tag/attribute list holding unit.

1 10. A structured document compressing apparatus  
2 according to claim 6, further comprising:  
3 a tag/attribute list holding unit for holding  
4 a tag/attribute list in which tags and an attribute name  
5 are listed in a predetermined order for the definition  
6 of a predetermined data structure;  
7 a tag/attribute rearranging unit for rearranging  
8 tags and an attribute in said structured document to be  
9 compressed, in the predetermined order according to the  
10 tag/attribute list held in said tag/attribute list  
11 holding unit; and  
12 an omitted tag/attribute supplementing unit for  
13 supplementing a tag and/or an attribute omitted in said  
14 structured document according to the tag/attribute list  
15 held in said tag/attribute list holding unit.

1 11. A method for compressing a plurality of  
2 structured documents having a common data structure, said  
3 method comprising the steps of:

4           obtaining a single tag list, common to said plural  
5   structured documents, that lists tags in the order of  
6   appearance;  
7           generating a plurality of compressed documents  
8   in which tags in individual said plural structured  
9   documents are replaced with predetermined delimiter  
10   codes; and  
11           outputting the single tag list and the plurality  
12   of compressed documents generated from said plural  
13   structured documents, in correspondence with one  
14   another.

1   12.       A method for compressing a structured document,  
2   said method comprising the steps of:  
3           detecting each tag in said structured document;  
4   and  
5           replacing said tag with a predetermined  
6   delimiter code.

1   13.       A method for compressing a structured document,  
2   said method comprising the steps of:  
3           extracting a subdocument, which is a region  
4   sandwiched between a start tag and an end tag that have  
5   a predetermined element name, from said structured  
6   document;  
7           detecting each tag in said subdocument; and  
8           replacing said detected tag with a predetermined

9 delimiter code.

1 14. A computer readable record medium which stores  
2 a structured document compressing program for  
3 instructing a computer to execute a function of  
4 compressing a plurality of structured documents having  
5 a common data structure, wherein said structured document  
6 compressing program instructs the computer to function  
7 as:

8 a tag list obtaining unit for obtaining a single  
9 tag list, common to said plural structured documents,  
10 that lists tags in the order of appearance;

11 a structured document compressing unit for  
12 generating a plurality of compressed documents in which  
13 tags in individual said plural structured documents are  
14 replaced with predetermined delimiter codes; and

15 an outputting unit for outputting said single  
16 tag list, which is obtained by said tag list obtaining  
17 unit, and also said plurality of compressed documents,  
18 which are generated individually from said plural  
19 structured documents by said structured document  
20 compressing unit, in correspondence with one another.

1 15. A computer readable record medium which stores  
2 a structured document compressing program for  
3 instructing a computer to execute a function of  
4 compressing a structured document, wherein said

5 structured document compressing program instructs the  
6 computer to function as:  
7 a tag detecting unit for detecting each tag in  
8 said structured document; and  
9 a tag replacement unit for replacing said tag,  
10 detected by said tag detecting unit, with a predetermined  
11 delimiter code.

1 16. A computer readable record medium which stores  
2 a structured document compressing program for  
3 instructing a computer to execute a function of  
4 compressing a structured document, wherein said  
5 structured document compressing program instructs the  
6 computer to function as:  
7 a subdocument extracting unit for extracting a  
8 subdocument, which is a region sandwiched between a start  
9 tag and an end tag that have a predetermined element name,  
10 from said structured document;  
11 a tag detecting unit for detecting each tag in  
12 said subdocument extracted by said subdocument  
13 extracting unit; and  
14 a tag replacement unit for replacing said tag,  
15 detected by said tag detecting unit, with a predetermined  
16 delimiter code.

1 17. An apparatus for decompressing a plurality of  
2 compressed documents, which are generated by replacing



3 tags in a plurality of original structured documents  
4 having a common data structure with predetermined  
5 delimiter codes, on the basis of a tag list in which tags  
6 in said plural original structured documents are listed  
7 in the order of appearance, said apparatus comprising:  
8       a duplicating unit for expanding/duplicating a  
9 data structure corresponding to said tag list, as a  
10 duplicated data structure, on a memory; and  
11       a writing unit for writing element contents of  
12 each of said compressed documents into predetermined  
13 regions of said duplicated data structure extended on  
14 said memory, in accordance with a correspondence between  
15 a position of a tag in said duplicated data structure  
16 and a position of the predetermined delimiter code in  
17 each of said compressed documents.

1 18.       An apparatus for decompressing a compressed  
2 document generated by replacing tags in an original  
3 structured document with predetermined delimiter codes,  
4 said apparatus comprising:  
5       a tag list holding unit for holding a tag list  
6 in which tags in said structured document are listed in  
7 the order of appearance;  
8       a delimiter code detecting unit for detecting  
9 each of the predetermined delimiter codes in said  
10 compressed document; and  
11       a tag restoring unit for replacing the

12 predetermined delimiter code, detected by said delimiter  
13 code detecting unit, with a corresponding tag on said  
14 tag list, in accordance with a correspondence between  
15 a position of the tag in said tag list and a position  
16 of the predetermined delimiter code detected by said  
17 delimiter code detecting unit.

1 19. An apparatus for decompressing a compressed  
2 document generated by replacing tags in a subdocument,  
3 which is a region, in an original structured document,  
4 sandwiched between a start tag and an end tag that have  
5 a predetermined element name, with predetermined  
6 delimiter codes, said apparatus comprising:

7 a tag list holding unit for holding a tag list  
8 in which tags in said subdocument are listed in the order  
9 of appearance;

10 a subdocument extracting unit for extracting  
11 said subdocument from said compressed document;

12 a delimiter code detecting unit for detecting  
13 each of the predetermined delimiter codes in said  
14 subdocument extracted by said subdocument extracting  
15 unit; and

16 a tag restoring unit for replacing the  
17 predetermined delimiter code, detected by said delimiter  
18 code detecting unit, with a corresponding tag on said  
19 tag list, in accordance with a correspondence between  
20 a position of the tag in said tag list and a position

21 of the predetermined delimiter code detected by said  
22 delimiter code detecting unit.

1 20. A structured document decompressing apparatus  
2 according to claim 18, wherein if an attribute inside  
3 an attribute-bearing tag in said original structured  
4 document is replaced with a set of an attribute value  
5 and a predetermined delimiter code in said compressed  
6 document, said apparatus further comprises:  
7 an attribute list holding unit for holding an  
8 attributelistinwhichattributenamesinsaidcompressed  
9 document are listed in the order of appearance;  
10 an attribute-bearing-tag discriminating unit  
11 for discriminating whether or not a given tag to be  
12 restored by said tag restoring unit is an  
13 attribute-bearing tag; and  
14 an attribute-bearing-tag restoring unit for  
15 restoring said attribute-bearing tag discriminated by  
16 said attribute-bearing-tag discriminating unit, in  
17 accordance with a correspondence between an attribute  
18 value for said attribute-bearing tag and an attribute  
19 name in said attribute list.

1 21. A structured document decompressing apparatus  
2 according to claim 19, wherein if an attribute inside  
3 an attribute-bearing tag in said original structured  
4 document is replaced with a set of an attribute value

5 and a predetermined delimiter code in said compressed  
6 document, said apparatus further comprises:

7 an attribute list holding unit for holding an  
8 attribute list in which attribute names in said compressed  
9 document are listed in the order of appearance;

10 an attribute-bearing-tag discriminating unit  
11 for discriminating whether or not a given tag to be  
12 restored by said tag restoring unit is an  
13 attribute-bearing tag; and

14 an attribute-bearing-tag restoring unit for  
15 restoring said attribute-bearing tag discriminated by  
16 said attribute-bearing-tag discriminating unit, in  
17 accordance with a correspondence between an attribute  
18 value for said attribute-bearing tag and an attribute  
19 name in said attribute list.

1 22. A method for decompressing a plurality of  
2 compressed documents, which is generated by replacing  
3 tags in a plurality of original structured documents  
4 having a common data structure with predetermined  
5 delimiter codes, on the basis of a tag list in which tags  
6 in said plural original structured documents are listed  
7 in the order of appearance, said method comprising the  
8 steps of:

9 expanding/duplicating a data structure  
10 corresponding to said tag list, as a duplicated data  
11 structure, on a memory; and

12           writing element contents of each of said  
13 compressed documents into predetermined regions of said  
14 duplicated data structure extended on said memory, in  
15 accordance with a correspondence between a position of  
16 a tag in said duplicated data structure and a position  
17 of the predetermined delimiter code in each of said  
18 compressed documents.

1   23.       A method for decompressing a compressed document  
2 generated by replacing tags in an original structured  
3 document with predetermined delimiter codes, said method  
4 comprising the steps of:  
5           holding a tag list in which tags in said structured  
6 document are listed in the order of appearance;  
7           detecting each of the predetermined delimiter  
8 codes in said compressed document; and  
9           replacing the detected predetermined delimiter  
10 code with a corresponding tag on said tag list, in  
11 accordance with a correspondence between a position of  
12 the detected predetermined delimiter code and a position  
13 of the tag in said tag list.

1   24.       A method for decompressing a compressed document  
2 generated by replacing tags in a subdocument, which is  
3 a region, in an original structured document, sandwiched  
4 between a start tag and an end tag that have a predetermined  
5 element name, with predetermined delimiter codes, said

6 method comprising the steps of:  
7           holding a tag list in which tags in said  
8 subdocument are listed in the order of appearance;  
9           extractingsaidsubdocument fromsaidcompressed  
10 document;  
11           detecting each of the predetermined delimiter  
12 codes in said extracted subdocument; and  
13           replacing the detected predetermined delimiter  
14 code with a corresponding tag on said tag list, in  
15 accordance with a correspondence between a position of  
16 the detected predetermined delimiter code and a position  
17 of the tag in said tag list.

1 25.       A computer readable record medium which stores  
2 a structured document decompressing program for  
3 instructing a computer to execute a function of  
4 decompressing a plurality of compressed documents  
5 generated by replacing tags, in a plurality of original  
6 structured documents having a common data structure, with  
7 predetermined delimiter codes on the basis of a tag list  
8 in which tags in said plural structured documents are  
9 listed in the order of appearance, wherein said structured  
10 document decompressing program instructs the computer  
11 to function as:

12           a duplicating unit for expanding/duplicating a  
13 data structure corresponding to said tag list, as a  
14 duplicated data structure, on a memory; and

15           a writing unit for writing element contents of  
16 each of said compressed documents into predetermined  
17 regions of said duplicated data structure extended on  
18 said memory, in accordance with a correspondence between  
19 a position of a tag in said duplicated data structure  
20 and a position of the predetermined delimiter code in  
21 each of said compressed documents.

1   26.       A computer readable record medium which stores  
2 a structured document decompressing program for  
3 instructing a computer to execute a function of  
4 decompressing a compressed document generated by  
5 replacing tags, in an original structured document, with  
6 predetermined delimiter codes, wherein said structured  
7 document decompressing program instructs the computer  
8 to function as:

9           a delimiter code detecting unit for detecting  
10 each of the predetermined delimiter codes in said  
11 compressed document; and

12           a tag restoring unit for replacing the  
13 predetermined delimiter code, detected by said delimiter  
14 code detecting unit, with a corresponding tag on a tag  
15 list in which tags in said structured document are listed  
16 in the order of appearance, in accordance with a  
17 correspondence between a position of the tag in said tag  
18 list and a position of the predetermined delimiter code  
19 detected by said delimiter code detecting unit.

1 27. A computer readable record medium which stores  
2 a structured document decompressing program for  
3 instructing a computer to execute a function of  
4 decompressing a compressed document generated by  
5 replacing tags in a subdocument, which is a region, in  
6 an original structured document, sandwiched between a  
7 start tag and an end tag that have a predetermined element  
8 name, with predetermined delimiter codes, wherein said  
9 structured document decompressing program instructs the  
10 computer to function as:

11 a subdocument extracting unit for extracting  
12 said subdocument from said compressed document;

13 a delimiter code detecting unit for detecting  
14 each of the predetermined delimiter codes in said  
15 subdocument extracted by said subdocument extracting  
16 unit; and

17 a tag restoring unit for replacing the  
18 predetermined delimiter code, detected by said delimiter  
19 code detecting unit, with a corresponding tag on a tag  
20 list in which tags in said subdocument are listed in the  
21 order of appearance, in accordance with a correspondence  
22 between a position of the tag in said tag list and a position  
23 of the predetermined delimiter code detected by said  
24 delimiter code detecting unit.

1 28. A structured document processing system for  
2 processing a plurality of structured documents having



3 a common data structure, comprising a structured document  
4 compressing apparatus for compressing said plurality of  
5 structured documents and a structured document  
6 decompressing apparatus for decompressing the data  
7 compressed by said structured document compressing  
8 apparatus, wherein  
9           said structured document compressing apparatus  
10 comprises:  
11                 a tag list obtaining unit for obtaining  
12 a single tag list, common to said plural structured  
13 documents, that lists tags, extracted from said plural  
14 structured documents, in the order of appearance;  
15                 a structured document compressing unit  
16 for generating a plurality of compressed documents in  
17 which tags in individual said structured documents are  
18 replaced with predetermined delimiter codes; and  
19                 an outputting unit for outputting said  
20 single tag list, which is obtained by said tag list  
21 obtaining unit, and also said plurality of compressed  
22 documents, which are generated individually from said  
23 plural structured documents by said structured document  
24 compressing unit, in correspondence with one another,  
25 and wherein  
26           said structured document decompressing unit  
27 comprises:  
28                 a duplicating unit for  
29 expanding/duplicating a data structure corresponding to

30 said tag list, as a duplicated data structure, on a memory;  
31 and  
32 a writing unit for writing element  
33 contents of each of said compressed documents into  
34 predetermined regions of said duplicated data structure  
35 extended on said memory, in accordance with a  
36 correspondence between a position of a tag in said  
37 duplicated data structure and a position of the  
38 predetermined delimiter code in each of said compressed  
39 documents.

1 29. A structured document processing system for  
2 processing a structured document, comprising a  
3 structured document compressing apparatus for  
4 compressing said structured document and a structured  
5 document decompressing apparatus for decompressing the  
6 data compressed by said structured document compressing  
7 apparatus, wherein  
8 said structured document compressing apparatus  
9 comprises:  
10 a tag detecting unit for detecting each  
11 tag in said structured document; and  
12 a tag replacement unit for replacing said  
13 tag, detected by said tag detecting unit, with a  
14 predetermined delimiter code, and wherein  
15 said structured document decompressing  
16 apparatus comprises:

17                   a tag list holding unit for holding a tag  
18 list in which tags in said structured document are listed  
19 in the order of appearance;

20                   a delimiter code detecting unit for  
21 detecting each of the predetermined delimiter codes in  
22 the data compressed by said structured document  
23 decompressing apparatus; and

24                   a tag restoring unit for replacing the  
25 predetermined delimiter code, detected by said delimiter  
26 code detecting unit, with a corresponding tag on said  
27 tag list, in accordance with a correspondence between  
28 a position of the tag in said tag list and a position  
29 of the predetermined delimiter code detected by said  
30 delimiter code detecting unit.

1   30.       A structured document processing system for  
2 processing a structured document, comprising a  
3 structured document compressing apparatus for  
4 compressing said structured document and a structured  
5 document decompressing apparatus for decompressing the  
6 data compressed by said structured document compressing  
7 apparatus, wherein

8                   said structured document compressing apparatus  
9 comprises:

10                   a first subdocument extracting unit for  
11 extracting a subdocument, which is a region sandwiched  
12 between a start tag and an end tag that have a predetermined

13 element name, from said structured document;  
14                   a tag detecting unit for detecting each  
15 tag in said subdocument extracted by said first  
16 subdocument extracting unit; and  
17                   a tag replacement unit for replacing said  
18 tag, detected by said tag detecting unit, with a  
19 predetermined delimiter code, and wherein  
20                   said structured document decompressing  
21 apparatus comprises:  
22                   a tag list holding unit for holding a tag  
23 list in which tags in said subdocument are listed in the  
24 order of appearance;  
25                   a second subdocument extracting unit for  
26 extracting said subdocument from the data compressed by  
27 said structured document compressing apparatus;  
28                   a delimiter code detecting unit for  
29 detecting each of the predetermined delimiter codes in  
30 said subdocument extracted by said second subdocument  
31 extracting unit; and  
32                   a tag restoring unit for replacing the  
33 predetermined delimiter code, detected by said delimiter  
34 code detecting unit, with a corresponding tag on said  
35 tag list, in accordance with a correspondence between  
36 a position of the tag in said tag list with a position  
37 of the predetermined delimiter code detected by said  
38 delimiter code detecting unit.

1 31. A structured document processing system  
2 according to claim 29, further comprising:  
3 a tag-list-group holding unit for holding a  
4 plurality of tag lists corresponding to data structures  
5 of structured documents that can possibly be processed;  
6 and  
7 a tag list managing unit for managing  
8 correspondence between compressed documents generated  
9 by said structured document compressing apparatus and  
10 said tag lists held in said tag-list-group holding unit.

1 32. A structured document processing system  
2 according to claim 30, further comprising:  
3 a tag-list-group holding unit for holding a  
4 plurality of tag lists corresponding to data structures  
5 of structured documents that can possibly be processed;  
6 and  
7 a tag list managing unit for managing  
8 correspondence between compressed documents generated  
9 by said structured document compressing apparatus and  
10 said tag lists held in said tag-list-group holding unit.

1 33. A structured document processing system  
2 according to claim 29, further comprising:  
3 a tag-list-group holding unit for holding a  
4 plurality of tag lists corresponding to data structures  
5 of structured documents that can possibly be processed;

6           a tag-list identification information adding  
7 unit for adding tag-list identification information,  
8 which identifies a tag list that corresponds to a  
9 compressed document generated by said structured  
10 document compressing apparatus, to said compressed  
11 document; and

12           a tag-list identification information obtaining  
13 unit for obtaining said tag-list identification  
14 information added to said compressed document,

15           said structured document decompressing  
16 apparatus decompressing said compressed document using  
17 said tag list that corresponds to said tag-list  
18 identification information obtained by said tag-list  
19 identification information obtaining unit.

1   34.       A structured document processing system  
2 according to claim 30, further comprising:

3           a tag-list-group holding unit for holding a  
4 plurality of tag lists corresponding to data structures  
5 of structured documents that can possibly be processed;

6           a tag-list identification information adding  
7 unit for adding tag-list identification information,  
8 which identifies a tag list that corresponds to a  
9 compressed document generated by said structured  
10 document compressing apparatus, to said compressed  
11 document; and

12           a tag-list identification information obtaining

13 unit for obtaining said tag-list identification  
14 information added to said compressed document,  
15 said structured document decompressing  
16 apparatus decompressing said compressed document using  
17 said tag list that corresponds to said tag-list  
18 identification information obtained by said tag-list  
19 identification information obtaining unit.

1 35. A structured document processing system  
2 according to claim 31, wherein said tag-list-group  
3 holding unit is provided on a management server, which  
4 is communicably connected with said structured document  
5 compressing apparatus and with said structured document  
6 decompressing apparatus via a network, and a tag list  
7 necessary for the processing is read from said  
8 tag-list-group holding unit on said management server.

1 36. A structured document processing system  
2 according to claim 32, wherein said tag-list-group  
3 holding unit is provided on a management server, which  
4 is communicably connected with said structured document  
5 compressing apparatus and with said structured document  
6 decompressing apparatus via a network, and a tag list  
7 necessary for the processing is read from said  
8 tag-list-group holding unit on said management server.

1 37. A structured document processing system

2 according to claim 33, wherein said tag-list-group  
3 holding unit is provided on a management server, which  
4 is communicably connected with said structured document  
5 compressing apparatus and with said structured document  
6 decompressing apparatus via a network, and a tag list  
7 necessary for the processing is read from said  
8 tag-list-group holding unit on said management server.

1 38. A structured document processing system  
2 according to claim 34, wherein said tag-list-group  
3 holding unit is provided on a management server, which  
4 is communicably connected with said structured document  
5 compressing apparatus and with said structured document  
6 decompressing apparatus via a network, and a tag list  
7 necessary for the processing is read from said  
8 tag-list-group holding unit on said management server.